



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/750,011	12/31/2003	Andrew S. Laucius	MS306413.1	9331
27195 7590 09/21/2007 AMIN. TUROCY & CALVIN, LLP 24TH FLOOR, NATIONAL CITY CENTER 1900 EAST NINTH STREET CLEVELAND, OH 44114			EXAMINER SCIACCA, SCOTT M	
			ART UNIT 2146	PAPER NUMBER
			NOTIFICATION DATE 09/21/2007	DELIVERY MODE ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

docket1@thepatentattorneys.com
hholmes@thepatentattorneys.com
osteuball@thepatentattorneys.com

Office Action Summary

Application No.

10/750,011

Applicant(s)

LAUCIUS ET AL.

Examiner

Scott M. Sciacca

Art Unit

2146

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 31 December 2003.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-28 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-28 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 31 December 2003 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. _____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | 5) <input type="checkbox"/> Notice of Informal Patent Application |
| 3) <input checked="" type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08)
Paper No(s)/Mail Date <u>4/19/2004</u> . | 6) <input type="checkbox"/> Other: _____ |

DETAILED ACTION

Claim Rejections - 35 USC § 102

1. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

2. Claims 26-28 are rejected under 35 U.S.C. 102(e) as being anticipated by Harik (US 7,213,198).

Regarding Claim 26, Harik teaches a system that facilitates increment web crawls comprising:

means for placing items with similar properties into respective chunks (*"In another embodiment, a computer implemented method of grouping hyperlinked documents is provided"* – See Col. 2, lines 29-30); and,

means for storing at least some of the properties associated with the respective chunk (*"Fig. 11 shows a hash table that can be utilized to organize the relationships between the web pages in Fig. 10"* – See Col. 8, lines 51-52).

Regarding Claim 27, Harik teaches the system of Claim 26, the items comprising information associated with a Uniform Resource Locator (*"The web pages typically include links in the form of uniform resource locators (URLs) that are a link to another web page"* – See Col. 4, lines 50-52).

Regarding Claim 28, Harik teaches the system of Claim 26, the items comprising at least one of an HTML file, a PDF file, a PS file, a PPT file, an XLS file and a DOC file (Fig. 18A shows several HTML documents in a clustered group of search results).

Claim Rejections - 35 USC § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

4. Claims 1-4, 6-14, 17 and 19-21 are rejected under 35 U.S.C. 103(a) as being unpatentable over Harik (US 7,213,198) in view of Agarwal et al. (US 2004/0225963).

Regarding Claim 1, Harik teaches a system that facilitates web crawls comprising (*"information about what web pages point to each web page can be collected and stored by a web crawler"* – See Col. 8, lines 18-20):

an indexer that places items with similar properties into respective chunks (*"In another embodiment, a computer implemented method of grouping hyperlinked documents is provided"* – See Col. 2, lines 29-30); and,

a chunk map that stores at least some of the properties associated with the respective chunk, the chunk map employed to facilitate an incremental web re-crawl (*"Fig. 11 shows a hash table that can be utilized to organize the relationships between the web pages in Fig. 10"* – See Col. 8, lines 51-52).

Harik does not explicitly teach a system that facilitates *incremental* (emphasis added) web crawls, however Agarwal does teach a system that facilitates incremental web crawls (*"In one embodiment of the invention, the search engine server 110 utilizes a crawler (not shown) to automatically find documents in the document repository 104 and update the search engine's records"* – See [0025]; *"It has been shown that an in-place, incremental crawler can improve the freshness of the inverted index"* – See [0005]).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to use an incremental web crawler. One would have been motivated to use an incremental web crawler in order to provide an effective method of keeping the collection of web documents as current as possible.

Regarding Claim 2, Harik in view of Agarwal teaches the system of Claim 1. Additionally, Harik teaches the items comprising information associated with a Uniform Resource Locator (*"The web pages typically include links in the form of uniform resource locators (URLs) that are a link to another web page"* – See Col. 4, lines 50-52).

Regarding Claim 3, Harik in view of Agarwal teaches the system of Claim 1. Additionally, Harik teaches the items comprising at least one of an HTML file, a PDF file, a PS file, a PPT file, an XLS file and a DOC file (Fig. 18A shows several HTML documents in a clustered group of search results).

Regarding Claim 4, Harik in view of Agarwal teaches the system of Claim 1. Additionally, Harik teaches the crawler being responsible for a specific set of Uniform

Art Unit: 2146

Resource Locators (Group the hyperlinked documents according to the forward links of the other hyperlinked documents (305) – See Fig. 5).

Regarding Claim 6, Harik in view of Agarwal teaches the system of Claim 1. Additionally, Agarwal teaches the system further comprising a master control process that serves as an interface between a crawler and a re-crawl controller (*"In one embodiment of the invention, the search engine server 110 utilizes a crawler (not shown) to automatically find documents in the document repository 104 and update the search engine's records"* – See [0025]).

Regarding Claim 7, Harik in view of Agarwal teaches the system of Claim 6. Additionally, Harik teaches the system wherein the master control process maintains a known chunks table that stores information for components of a system (*"Fig. 11 shows a hash table that can be utilized to organize the relationships between the web pages in Fig. 10"* – See Col. 8, lines 51-52).

Regarding Claims 8, Harik in view of Agarwal teaches the system of Claim 6. Additionally, Harik teaches the system wherein the master control process exposes an interface for communication with a component of the system (*"FIGS. 18A-8C show a web page example of search results being grouped by topic"* – See Col. 12, lines 58-59).

Regarding Claims 9 and 10, Harik in view of Agarwal teaches the system of Claim 8. Additionally, Harik teaches the interface returning a list of chunks the component should have and where to get the chunks as well as the interface returning a list of the chunks that should be actively served by the component (*"The query was for*

Art Unit: 2146

web pages related to "Saturn." As shown, a first group 1201 includes pages that are related to the Planet Saturn, a second group 1203 includes pages that are related to the Saturn car and a third group include pages that are related to the Sega Saturn game system. Each group graphically separated or sectioned off from the other groups" – See Col. 12, lines 59-65).

Regarding Claim 11, Harik in view of Agarwal teaches the system of Claim 8. Additionally, Harik teaches the interface returning a range of chunk identifiers to use in building a new chunk by the component (*"In order to generate hash table 801, a position for the originating document is identified and all the links from this originating are then stored at that position" – See Col. 9, lines 4-6).*

Regarding Claim 12, Harik in view of Agarwal teaches the system of Claim 8. Additionally, Agarwal teaches the interface causing an old chunk to be retired by the system (*"One solution known in the art for keeping document repositories up-to-date is to rebuild the index more frequently" – See [0006]).*

Regarding Claim 13, Harik in view of Agarwal teaches the system of Claim 6. Additionally, Harik teaches the master control process facilitating movement of chunks from one component to another component (*"The query was for web pages related to "Saturn." As shown, a first group 1201 includes pages that are related to the Planet Saturn, a second group 1203 includes pages that are related to the Saturn car and a third group include pages that are related to the Sega Saturn game system. Each group graphically separated or sectioned off from the other groups" – See Col. 12, lines 59-65).*

Regarding Claim 14, Harik in view of Agarwal teaches the system of Claim 13. Additionally, Agarwal teaches movement of chunks being based, at least in part, upon at least one of rebalancing index servers after one goes down, re-crawling pages previously crawled (*"One solution known in the art for keeping document repositories up-to-date is to rebuild the index more frequently"* – See [0006]), and, restoring a state of a crawler after it has crashed.

Regarding Claim 17, Harik in view of Agarwal teaches the system of Claim 1. Additionally, Harik teaches an index chunk that stores information associated with an index of at least some of the items (See Fig. 12A & 12B).

Regarding Claim 19, Harik teaches parsing a first chunk for uniform resource locators (*"In the description that follows, systems and methods consistent with the principles of the invention will be described in reference to embodiments that group hyperlinked documents (e.g., web pages)"* – See Col. 3, lines 56-59; *"The web pages typically include links in the form of uniform resource locators (URLs) that are a link to another web page, whether it is on the same server or a different one"* – See Col. 4, lines 50-53), but does not explicitly teach re-crawling the uniform resource locators and forming a second chunk based, at least in part, upon the re-crawled uniform resource locators. However, Agarwal does teach re-crawling the uniform resource locators and forming a second chunk based, at least in part, upon the re-crawled uniform resource locators (*"Since web documents change frequently, keeping inverted indexes up-to-date is crucial in making the most recently indexed documents searchable. A crawler (also referred to as a spider) is a program that collects web documents to be indexed"* – See

Art Unit: 2146

[0005]; *“One solution known in the art for keeping document repositories up-to-date is to rebuild the index more frequently”* – See [0006]). It would have been obvious to one of ordinary skill in the art at the time the invention was made to re-crawl the web pages in a chunk and form a second chunk with the updated information. Motivation for doing so would be to have the freshest information available contained in the chunk as well as omit from the chunk any URLs that are no longer active (i.e., no longer exist).

Regarding Claim 20, Harik in view of Agarwal teaches the method of Claim 19. Additionally, Agarwal teaches the method comprising at least one of the following acts: determining whether any chunks are to be retired; moving the first chunk; and, destroying the first chunk (*“One solution known in the art for keeping document repositories up-to-date is to rebuild the index more frequently”* – See [0006]).

Regarding Claim 21, Harik in view of Agarwal teaches the method of Claim 19. Additionally, Harik teaches one or more computer readable media having stored thereon computer executable instructions for carrying out the method (*“FIG. 1 illustrates an example of a computer system that can be used to execute the software of an embodiment of the invention”* – See Col. 4, lines 4-6; *“Cabinet 7 houses a CD-ROM drive 13, system memory and a hard drive (see FIG. 2) which can be utilized to store and retrieve software programs incorporating computer code that implements the invention, data for use with the invention, and the like”* – See Col. 4, lines 9-13).

Art Unit: 2146

5. Claim 5 is rejected under 35 U.S.C. 103(a) as being unpatentable over Harik (US 7,213,198) in view of Agarwal et al. (US 2004/0225963) and further in view of Eichstaedt et al. (US 6,182,085).

Regarding Claim 5, Harik in view of Agarwal teaches the system of Claim 1. Additionally, Harik teaches a master control process that can modify the chunk map (*"Hash table 801 is used here in order to more efficiently group the links together"* – See Col. 8, lines 62-63; *"In order to generate hash table 801, a position for the originating document is identified and all the links from this originating are then stored at that position"* – See Col. 9, lines 4-6) but does not explicitly teach the system further comprising a master control process that can modify the chunk map to facilitate load balancing amongst a plurality of crawlers. However, Eichstaedt does teach a master control process that can modify the chunk map to facilitate load balancing amongst a plurality of crawlers (*"A distributed collection of web-crawlers to gather information over a large portion of the cyberspace. These crawlers share the overall crawling through a cyberspace partition scheme. They also collaborate with each other through load balancing to maximally utilize the computing resources of each of the crawlers. The invention takes advantage of the hierarchical nature of the cyberspace namespace and uses the syntactic components of the URL structure as the main vehicle for dividing and assigning crawling workload to individual crawler"* – See Abstract). It would have been obvious to one of ordinary skill in the art at the time the invention was made to have a control process that can facilitate load balancing amongst a plurality of crawlers. One

Art Unit: 2146

would have been motivated to do so in order to maximally utilize the computing resources of each of the crawlers.

6. Claims 15-16 are rejected under 35 U.S.C. 103(a) as being unpatentable over Harik (US 7,213,198) in view of Agarwal et al. (US 2004/0225963) and further in view of Najork et al. (US 6,263,364).

Regarding Claim 15, Harik in view of Agarwal teaches the system of Claim 1, but fails to explicitly teach the system further comprising a re-crawl component that employs the chunk map to determine which chunks, if any, to re-crawl at a particular time.

However, Najork does teach a re-crawl component that employs the chunk map to determine which chunks, if any, to re-crawl at a particular time (*"the host component of the document's URL; for example, documents from certain web sites known to the web crawler may be assigned a high or low download priority based on knowledge of how often documents at those web sites are updated"* – See Col. 11, lines 63-67). It would have been obvious to one of ordinary skill in the art at the time the invention was made to include a re-crawl component that determines which chunks to re-crawl. Motivation for doing so would be to keep the most important chunks up to date and give them greater priority for re-crawling over chunks that are of less importance.

Regarding Claim 16, Harik in view of Agarwal and further in view of Najork teaches the system of Claim 15. Additionally, Najork teaches the determination of which chunks to re-crawl, if any, being further based, at least in part, upon at least one of average time between change and average importance of documents comprising a

Art Unit: 2146

particular chunk (*"the host component of the document's URL; for example, documents from certain web sites known to the web crawler may be assigned a high or low download priority based on knowledge of how often documents at those web sites are updated"* – See Col. 11, lines 63-67).

7. Claim 18 is rejected under 35 U.S.C. 103(a) as being unpatentable over Harik (US 7,213,198) in view of Agarwal et al. (US 2004/0225963) and further in view of Acharaya et al. (US 2007/0094255).

Regarding Claim 18, Harik in view of Agarwal teaches the system of Claim 1, but does not explicitly teach a ranking chunk that stores a static rank associated with an index chunk. However, Acharaya does teach a ranking chunk that stores a static rank associated with an index chunk (*"Ranking component 330 may assign a ranking score (also called simply a "score" herein) to one or more documents in document corpus 340"* – See [0038]). It would have been obvious to one of ordinary skill in the art at the time the invention was made to store a rank associated with an index chunk. Motivation for doing so would be to improve search results generated in connection with a search query.

8. Claims 22-23 are rejected under 35 U.S.C. 103(a) as being unpatentable over Harik (US 7,213,198) in view of Najork et al. (US 6,263,364).

Regarding Claim 22, Harik teaches accessing a chunk map containing properties associated with respective chunks of data as a result of one or more web crawls (*"Fig.*

Art Unit: 2146

11 shows a hash table that can be utilized to organize the relationships between the web pages in Fig. 10” – See Col. 8, lines 51-52), but does not explicitly teach periodically determining, based on the properties in the chunk map, whether to re-crawl one or more of the chunks of data. However, Najork does teach periodically determining, based on the properties in the chunk map, whether to re-crawl one or more of the chunks of data (“the host component of the document's URL; for example, documents from certain web sites known to the web crawler may be assigned a high or low download priority based on knowledge of how often documents at those web sites are updated” – See Col. 11, lines 63-67). It would have been obvious to one of ordinary skill in the art at the time the invention was made to determine based on properties in the chunk map whether to re-crawl one or more chunks of data. Motivation for doing so is the same as that which was given with regard to Claim 15 above.

Regarding Claim 23, Harik in view of Najork teaches the method of Claim 22. Additionally, Najork teaches the period determination being based, at least in part, upon, at least one of average time between change and average importance of documents comprising a particular chunk (*“the host component of the document's URL; for example, documents from certain web sites known to the web crawler may be assigned a high or low download priority based on knowledge of how often documents at those web sites are updated” – See Col. 11, lines 63-67).*

9. Claims 24-25 are rejected under 35 U.S.C. 103(a) as being unpatentable over Harik (US 7,213,198) in view of Kosiba et al. (US 2003/0221014).

Regarding Claim 24, Harik teaches a data packet transmitted between two or more computer components that facilitates document re-crawl (*"information about what web pages point to each web page can be collected and stored by a web crawler"* – See Col. 8, lines 18-20; *"In another embodiment, a computer implemented method of grouping hyperlinked documents is provided"* – See Col. 2, lines 29-30), but fails to explicitly teach the data packet comprising a chunk header that includes metadata associated with the data packet, an offset section that provides offset information associated with document files, and the document files that include content found on the Internet. However, Kosiba does teach the data packet comprising a chunk header that includes metadata associated with the data packet, an offset section that provides offset information associated with document files, and the document files (*"Each download packet corresponds to a portion of a single component of the original file (i.e. a piece of an mdat atom). Each download packet may contain a media header containing information pertinent to the specific piece of data being downloaded including the components unique identifier, the size of the data in the current packet, and the offset within the file where this data may be placed"* – See [0114]).

Regarding Claim 25, Harik in view of Kosiba teaches the data packet of Claim 24. Additionally, Harik teaches at least one of the document files comprising at least one of an HTML file, a PDF file, a PS file, a PPT file, an XLS file and a DOC file (Fig. 18A shows several HTML documents in a clustered group of search results).

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Scott M. Sciacca whose telephone number is (571) 270-1919. The examiner can normally be reached on Monday thru Friday, 7:30 A.M. - 5:00 P.M. EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Jeff Pwu can be reached on (571) 272-6798. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

SS



JEFFREY PWU
SUPERVISORY PATENT EXAMINER